2022

Aggregate data yield biased estimates of voter preferences

Corey Lang

Shana Pearson-Merkowitz

Forthcoming, Journal of Environmental Economics and Management

AGGREGATE DATA YIELD BIASED ESTIMATES OF VOTER PREFERENCES

Corey Lang* and Shanna Pearson-Merkowitz

December 6, 2021

Abstract
Voter preferences and valuation of public goods are often estimated using aggregated votes matched with Census data at the same spatial scale. However, this method may yield biased estimates for two reasons we examine in this paper: using Census data ignores the selection process of who votes, and relying on comparisons between aggregated units makes models susceptible to omitted variable bias. To assess bias, we use both Monte Carlo simulation and a case study regarding a statewide environmental bond referendum for which we have collected aggregate data and individual exit poll data. Our results confirm the two sources of bias and show that aggregate model regression coefficients can be incorrect in magnitude and even sign. We conclude that using aggregate data will likely lead to incorrect assessment of valuation and distributional impacts of public good provision.

Keywords: valuation, voting, referendum, ecological fallacy, aggregation bias
JEL codes: C15, C43, D72, H41, Q51

---

*Corresponding author. Address: University of Rhode Island, Kingston, RI, 02881. Email: clang@uri.edu.
For useful feedback, we thank Christian Vossler, Sarah Jacobson, Todd Guilfoos, Mark Agerton, Soren Anderson, Jesse Burkhardt, Kathy Baylis, Laura Grant, Yusuke Kuwayama, Lucija Muehlenbachs, two anonymous referees, and conference/seminar participants at W4133, RFF, and UC Davis. Author contributions: CL conceived of research, collected data, analyzed data, and wrote paper; SPM collected data and provided feedback. This work was supported by the University of Rhode Island Coastal Institute and the USDA National Institute of Food and Agriculture, Agricultural and Food Research Initiative Competitive Program, Agriculture Economics and Rural Communities, grant number 2015-67024-22937. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1 INTRODUCTION

Direct democracy plays a critical role in shaping state laws and policies and setting the level of local public goods (Matsusaka 2005). In recent years, voters have judged ballot questions related to gay marriage, marijuana legalization, Medicaid, and gun rights, and have voted on bond and tax referendums worth billions of dollars for environmental protection, education, and infrastructure, among others.

Many papers have examined determinants of referendum outcomes using publicly available, aggregated votes to estimate voter preferences and valuation of public goods. Environmental applications are most common in areas such as land conservation (Deacon and Shapiro 1975, Kotchen and Powers 2006, Banzhaf et al. 2010), carbon mitigation (Holian and Kahn 2015, Anderson et al. 2019), river restoration (Deacon and Schläpfer 2010), and general environmental quality (Kahn and Matsusaka 1997, Burkhardt and Chan 2017). Additional topics studied include tax restrictions (Cutler et al. 1999), school finance (Brunner and Ross 2010), and Medicaid expansion (Matsa and Miller 2018), among others. Official individual vote choices are unavailable because of the right to cast a secret ballot, but official aggregated votes by precinct, town, or county are widely available. The method of these papers always involves matching aggregate votes to Census demographics at the same spatial scale, and then regressing aggregate vote approval on Census demographics. These descriptive regression results are interpreted as voter preferences by demographic group, and if cost is included in the model the results can estimate willingness to pay. Table A1 in the online appendix lists 49 known papers that follow this framework.

The purpose of this paper is methodological: do regression results using aggregate data yield unbiased estimates of voter preferences on specific referendums and valuation of public goods more generally? While authors in this literature sometimes acknowledge that aggregation bias could be an issue, they never dwell on it and always assume it is minimal.[1] However, we

[1] Some papers (e.g., Kotchen and Powers 2006) argue that aggregation bias is not problem by citing Fischel (1979), who surveys voters in eight, rural New Hampshire towns about a referendum on a paper mill. However, the basis for this claim is that Fischel finds mean approval is similar between a survey and the official vote, and thus this is more about survey validity than unbiased regression estimates. Alternatively, some papers claim only to be interested in an aggregate level, i.e., precinct or town level determinants of approval (e.g., Banzhaf et al. 2010). However, omitted variables will still bias estimates in this case. Rarely do papers discuss the selection problem of voter participation and that Census population data does not represent the voting population well. Burkhardt and Chan (2017) argue that “a referendum is typically posed alongside other referenda as well as elections for public office. Therefore, it is unlikely that there is a strong correlation between turnout and voting outcomes for the referendum of interest, as many other factors drive voters’ participation decision.” Essentially, they argue that there is no additional

find that aggregate model coefficients can be incorrect in magnitude and even sign, and can lead to incorrect assessment of valuation and distributional impacts of public good provision. We build on prior literature that has identified two sources of bias, but which have not been combined nor applied to a valuation setting. First, there may be omitted variables correlated with demographic characteristics and approval, and we focus particularly on spatially referenced omitted variables, which is an expression of ecological fallacy (Robinson 1950). People sort across neighborhoods, and people vote where they live, which means there are likely unobservable variables correlated with observable demographics and voter preferences, similar to findings in the housing market (Kuminoff et al. 2010). Because aggregate models rely entirely on spatial variation, or between-precinct variation, they are more susceptible to omitted spatial variable bias than models using individual data, which use both within- and between-precinct variation to estimate coefficients. The second source of bias is that aggregate Census population data mismeasure the voting population (McDonald and Popkin 2001). The choice to vote is a selection problem, and as a result voters are observably different than non-voters, with income, age, education, race, and ethnicity all correlated with who shows up to the polls (Leighley and Nagler 2013). Using Census population characteristics to proxy for the voting population ignores the selection problem and introduces systematic (non-random) measurement error in the independent variables.

We proceed on two fronts to understand the bias of estimating voter preferences using aggregate data. First, we develop a Monte Carlo simulation analysis that matches the salient structure of individual voting decisions and aggregated precinct records, and compare aggregate regression results to truth. Consistent with the problems identified above, the data generating process allows for selection in who votes, the presence of spatially omitted variables, and varying proportions of within- versus between-precinct variation in demographics. Unsurprisingly, spatial omitted variables cause bias; coefficients are downward (upward) biased when correlation is negative (positive). In terms of selection bias, when a group is disproportionately less likely to vote, the aggregate model regression coefficient for that group

selection issue for a given referendum, which may well be true, but obfuscates the larger point that using Census data ignores voter participation selection and does not accurately represent the socioeconomic characteristics of voters. Holian and Kahn (2015) actually analyze both individual and aggregate voter data and find substantial differences in coefficients between the two models. However, they do not highlight nor seek to explain these differences.

will be attenuated. If a group is more likely to vote, the coefficient will be amplified. When both sources of bias are present, the coefficient bias cannot be signed. In addition, we also model an exit poll that samples individual voters at a small subset of precincts. In every case, the exit poll outperforms the aggregate model because 1) it controls for voter selection by only including voters in the sample and 2) estimates coefficients using individual, within-precinct variation that mitigates or eliminates bias from spatial omitted variables. Further, unbiasedness holds even if participation in the exit poll is endogenous.

The second thrust of analysis is a case study of a statewide bond referendum for environmental spending that was held in Rhode Island in November 2016. We built precinct-level, aggregate data by matching precinct approval to socioeconomic information from the American Community Survey, which mirrors data construction in prior studies. Given the favorable performance of exit polls in the Monte Carlo analysis, we undertook an extensive exit poll that we use to benchmark the aggregate data regression results. We enlisted 80 undergraduate and graduate student volunteers who surveyed at 37 sample precinct locations and collected over 2,000 surveys as voters left the polls. Simple means comparisons between the two datasets suggest significant differences between the voting and general population, particularly among education and income groups, and in directions mostly consistent with the findings of Leighley and Nagler (2013). For example, while 37.2% of Rhode Island's adult population has a high school degree or less education, only 13.3% of our exit poll sample was at that education level. In addition, we estimate a large degree of within-precinct variation in referendum approval and socioeconomic variables. On average, about 95% of variation is within-precinct, meaning that aggregate models are using only 5% of total variation to estimate coefficients.

We estimate identical models of voter preferences for the bond referendum using the aggregate data and the individual exit poll data, regressing referendum approval on presidential vote and socioeconomic characteristics (age, education, income, and race). We find large and statistically significant differences between the two models. Coefficients on presidential vote are substantially different: the aggregate model's coefficient on Voted for Clinton is 32% larger than the individual model, and the aggregate model rejects the individual model's point estimate with over 99% confidence. Further, coefficients on voting for third party candidates are different by an order of magnitude. Coefficients on socioeconomic characteristics from the two models can differ in magnitude, significance, and sign, and present different pictures of which types of voters support the referendum. For example, the aggregate model indicates that those with education levels of high school degree or less and college degree are most likely to support the referendum, whereas the individual model suggests it is those with a graduate degree that have the strongest support. We estimate additional models that support the two sources of bias being present in the aggregate model. First, we are able to find suggestive evidence of the presence of spatial omitted variable bias by including additional spatial controls and spatial fixed effects. Second, we develop a procedure with the exit poll data that simulates ignoring voter selection, and we find evidence that this induced mismeasurement can explain part of the discrepancy between the aggregate and individual voter preference estimates.

Lastly, in the spirit of Burkhardt and Chan (2017), we estimate willingness to pay (WTP) for the environmental referendum and the distribution of WTP across demographic groups. We find that estimated WTP is over 2.5 times larger using the aggregate model than the individual model. In terms of distributional impacts, sometimes the aggregate and individual results agree about which group wins and which loses, but the magnitudes of disparities can be substantially different. However, in the cases of educational attainment and race/ethnicity the models disagree about the direction of disparate impacts. While with different referendums or different states, the direction of bias may be different, these findings demonstrate that estimated valuation derived from aggregate voting data are likely to lead to incorrect inferences about overall benefits and the distributional impacts of a given policy or proposal.

This paper significantly advances the literature on empirical estimation of voter preferences and valuation through real-world referendum outcomes by demonstrating that aggregate data yield biased estimates and why. The underlying causes of the bias have been studied separately before. Issues regarding voter turnout and voter selection are well studied (Gerber and Green 2000, McDonald and Popkin 2001, Hajnal and Lewis 2003, Leighley and Nagler 2013, DellaVigna et al. 2016), though these papers do not examine preferences, only the decision to show up at the polls. Ecological fallacy or aggregation bias stemming from spatial omitted variables has been studied across many disciplines including political science, epidemiology, statistics, and economics (Robinson 1950, King 1997, Firebaugh 1978, Piantadosi et al. 1988, Greenland and Morgenstern 1989, Gotway and Young 2002, Banzhaf et al. 2019).[2]

[2] The work in political science on this topic (e.g., King 1997) focuses on cross-level inference, such as estimating the number of registered Black voters given the number of registered voters and the number of Black people.

However, this is the first paper to recognize both of these issues as possible sources of bias when estimating voter valuation using aggregate data. The integration of concepts is critical because we find that that the direction of bias cannot even be signed when both sources are present. Given the prevalence of these applications and the importance of understanding distributional impacts of policy, this is a critical methodological contribution.

2 MONTE CARLO SIMULATIONS

2.1 Description of the Data Generating Process

We design a data generating process (DGP) that matches the key elements of voting data to examine how ignoring voter selection or spatial omitted variables can lead to biased regression estimates of voter preferences. There are a total of N people living in J precincts, with the population for precinct j denoted N_j. For simplicity, all precincts have the same number of people, $N_j = N/J$. Individuals are indexed by i and j. There is a single demographic characteristic, $x_{ij} \in \{0,1\}$, and the proportion of the population with $x_{ij} = 1$ is denoted μ. These two groups may not be evenly distributed between precincts, i.e., there may be sorting. The degree of sorting is measured by the within-precinct variation, which we label ω.[3] Larger values of ω imply less sorting (less segregation) and smaller values imply more sorting/segregation.

Not all people vote, and the non-random choice is dependent on x. d_{ij} is a binary variable that equals one if person i votes. We define the probability of voting for the $x = 0$ population as $\pi_1 = P(d_{ij} = 1 | x_{ij} = 0) \in (0,1)$ and the difference in probability of voting for the $x = 1$ population as $\theta_1 = P(d_{ij} = 1 | x_{ij} = 1) - P(d_{ij} = 1 | x_{ij} = 0)$. $\theta_1 > 0$ implies $x = 1$ people are more likely to vote than $x = 0$ people, and $\theta_1 < 0$ the opposite. While the total number of people in each precinct is equal, the number of voters may be different across precincts due to θ_1 and ω.

There is a referendum that voters vote on, and there is a latent variable, v_{ij}^*, that is interpreted as net utility from passage of the referendum. $v_{ij}^* = \alpha + \beta x_{ij} + u_j + \varepsilon_{ij}$, where u_j is

---

Robinson (1950) and Cho and Gaines (2004) are critical of the value of aggregate voting data in the context of cross-level inference, though Greiner and Quinn (2010) find that a combination of exit poll and aggregate data can be optimal under certain circumstances. Another example of aggregation bias research in political science is Gerber and Lewis (2004), who find that the influence of the median voter is dependent on within-district heterogeneity, which is only uncovered using individual data. These papers do not examine valuation.

[3] Formally, $\omega = 1 - R^2$, and R^2 comes from the regresion $x_{ij} = \alpha_j + \epsilon_{it}$, where α_j is a precinct fixed effect.

an unobserved variable that is correlated with precinct averages of x, defined by the function $u_j = \rho\bar{x}_j + \tau_j$, where $\tau_j \sim N(0,1)$, and $\varepsilon_{ij} \sim N(0, \sigma^2)$.[4] v_{ij}^* is unobserved, but the vote choice, or approval, v_{ij} is observed, with $v_{ij} = 1$ if $v_{ij}{}^* > 0$ and $v_{ij} = 0$ if $v_{ij}{}^* \leq 0$.

The goal of empirical analysis is to estimate the propensity of the $x = 1$ group to approve the referendum relative to the $x = 0$ group, which is the relative preferences of the groups. We define $\delta = E\left[v_{ij} \middle| x_{ij} = 1, u_j\right] - E\left[v_{ij} \middle| x_{ij} = 0, u_j\right]$ as the true measure of those relative preferences.

Regression analysis of aggregate data uses the average approval by precinct for people who voted, which we label $\bar{\bar{v}}_j = \sum_{i \in j}(v_{ij} | d_{ij} = 1) \,/\, \sum_{i \in j} d_{ij}$. Throughout this paper, two bars are used to denote averages of the voting population, and one bar is used for averages of the total population. At the aggregate level, the demographic characteristics of those who actually voted is not known, instead the average value of x for voters and non-voters, $\bar{x}_j = \sum_{i \in j} x_{ij} \,/ N_j$, is used as the key explanatory variable. The aggregate regression model is $\bar{\bar{v}}_j = \gamma_{agg} + \delta_{agg}\bar{x}_j + \varphi_j$ and has J observations.

To simulate an exit poll, a subset of precincts, $J^S \in J$, are randomly sampled, with the proportion of sampled precincts denoted p. Similar to the decision to vote, we allow the decision to participate in the exit poll (e_{ij}) to be non-random and dependent on x. The probability of participation for the $x = 0$ population is $\pi_2 = P\left(e_{ij} = 1 \middle| x_{ij} = 0, d_{ij} = 1, j \in J^S\right) \in (0,1)$ and the difference in probability of participation for the $x = 1$ population as $\theta_2 = P\left(e_{ij} = 1 \middle| x_{ij} = 1, d_{ij} = 1\right) - P\left(e_{ij} = 1 \middle| x_{ij} = 0, d_{ij} = 1\right)$. $\theta_2 > 0$ implies $x = 1$ people are more likely to participate in the exit poll than $x = 0$ people, and $\theta_2 < 0$ the opposite. The exit poll survey collects data on v_{ij} and x_{ij} from participants, which we assume are accurately reported, which is a fairly benign assumption given that exit polls are anonymous and confidential.[5] The basic regression model resulting from the exit poll data is $v_{ij} = \gamma_{ep} + \delta_{ep}x_{ij} + \varphi_{ij}$ and has $\sum e_{ij}$ observations.

[4] This formulation of ecological fallacy is consistent with modeling in other disciplines (e.g., Firebaugh 1978, Piantadosi et al. 1988), but instead of calling it omitted variable bias, a correlation between $\bar{x}_j$ and v_{ij}^* is typically called group effects.
[5] Further, census data are also self-reported, so any misreporting in exit polls will likely also be present in the aggregate data.

We simulate the DGP under various parameterizations 100 times each in order to assess the distribution of coefficient bias present in the two regression models. For each iteration, we estimate $\hat{\delta}_{agg}$ and $\hat{\delta}_{ep}$, and difference these coefficients from δ to calculate bias. All models are estimated using OLS, though logit analysis yields qualitatively identical findings.

Important Simulation parameters	
ρ	Sets correlation between spatial omitted variable and precinct mean demographics
ω	Proportion of variation in x that is within-precinct
μ	Proportion of the population in $x = 1$ group
θ_1	Differential rate of voting for $x = 1$ group
θ_2	Differential rate of exit poll participation for $x = 1$ group

For all results presented below, the following parameter values are used: $N = 250{,}000$, $J = 500$, $N_j = 500$, $p = 0.1$, $\pi_1 = 0.5$, $\pi_2 = 0.1$, $\beta = 1$, and $\sigma = 2$. α adjusts across parameterizations in order to ensure that $E\left[P\left(v_{ij}{}^{*} > 0\right)\right] = 0.5$. The critical parameters that we vary are ρ, μ, ω, θ_1, and θ_2, which appear in the text box for reference.[6]

2.2 Assessing the role of spatial omitted variables

The critical parameters for understanding spatial omitted variable bias are ρ, which is the parameter that sets $Corr\left(\overline{x_j}, u_j\right)$, and ω, the proportion of within variation. We set $\mu = 0.2$, but other values produce similar results, and we set $\theta_1 = 0$ and $\theta_2 = 0$ in order to focus solely on the problem of omitted variables for now.

Figure 1 displays two plots of estimated biases for the two estimators. The left panel contains simulations with a large proportion of within-precinct variation, $\omega \sim 95\%$, and the right panel contains less, $\omega \sim 80\%$. These values were chosen based on the ranges observed in the exit poll data (explained below in Section 3.3). The vertical axis of each plot measures coefficient bias and the horizontal axis is different values of ρ, ranging from large negative to large positive.

For the aggregate model, the results are similar across plots; downward bias occurs when there is negative correlation and upward bias occurs when there is positive correlation. Further, the magnitude of bias grows with the magnitude of correlation. This finding is hardly novel, as it is consistent with standard omitted variable bias, in which bias is determined by the sign and strength of correlation and the coefficient on the unobserved variable. More novel, however, and specific to aggregation, we see that as the within-precinct variation decreases (moving from the

[6] Simulation code is available by request.

left panel to the right), the variance of estimates decreases, so in essence the aggregate model is more consistently biased, though the magnitude of bias is unchanged. This occurs because as between-precinct variation increases, the more variation the aggregate model is able to use for estimation and the more precise the estimates become.

For the individual exit poll model, in addition to ρ, bias depends on the amount of within variation. In the left panel, there is no visual evidence of bias, though it does exist, but is very small in magnitude, with the sign of bias following that of the aggregate model. However in the right panel, the magnitude of bias increases, but is still four to five times smaller than the bias of the aggregate model. The intuition behind these results is that the exit poll model is able to use within-precinct variation for estimation, and this variation is less correlated with the unobserved variable than the between-precinct variation, $\left|Corr\left(\overline{x_j}, u_j\right)\right| > \left|Corr\left(x_{ij}, u_j\right)\right| > 0$. As the within-precinct variation declines, the $\left|Corr\left(x_{ij}, u_j\right)\right|$ increases and the exit poll bias increases. When all variation is between-precinct, meaning absolute segregation, then aggregate and individual model have equal bias.[7]

2.3 Assessing the role of voter selection

The aggregate model ignores voting participation selection and assumes that all people are equally likely to vote, and thus population demographic averages equal voter demographic averages. The critical parameters for understanding how this assumption may lead to bias are μ, the proportion of the population with $x = 1$, and θ_1, the differential rate of voting for those with $x = 1$. For all iterations, we set $\rho = 0$, $\tau_j = 0$, and $\theta_2 = 0$ to isolate the problem of mismeasurement. We set $\omega \sim 80\%$, but other values produce similar results.

Figure 2 presents two panels: the left sets $\theta_1 = -0.2$ and the right sets $\theta_1 = 0.2$. The vertical axis of each plot measures coefficient bias and the horizontal axis is different values of μ, ranging from 0.2 to 0.8.

For the aggregate model, the left panel shows that when the $x = 1$ group is less likely to vote, $\hat{\delta}_{agg}$ is biased downwards when the $x = 1$ group is less than half of the population and is

[7] While a spatial omitted variable is the focus here (consistent with prior ecological fallacy literature), there may also be an individual-level omitted variable. The online appendix explores this and finds that the aggregate model and exit poll model are equally biased in this case. In real applications, votes for president and a large number of socioeconomic variables are included in the regression model, thus individual-level omitted variables seems unlikely.

biased upwards when the $x = 1$ group is more than half of the population. The right panel shows the opposite pattern: when the $x = 1$ group is more likely to vote, $\hat{\delta}_{agg}$ is biased upwards (downwards) when the $x = 1$ group is less (more) than half of the population. The sign of bias is dependent on the sign of δ. Since δ is positive in these results, downward bias is attenuation bias and upward bias is amplification bias. Figure A2 in the online appendix presents results for a negative δ, and the sign of bias is flipped from what appears in Figure 2.

The intuition behind these results stems from understanding how variance changes with the systematic measurement error created by ignoring selection. For the aggregate model, $\hat{\delta}_{agg} = cov(\bar{x}_j, \bar{\bar{v}}_j)/var(\bar{x}_j)$. The true proportion of $x = 1$ population voting is denoted $\bar{\bar{x}}_j$, and thus we can define the aggregate regression coefficient without a mismeasured voting population $\hat{\delta}'_{agg} = cov(\bar{\bar{x}}_j, \bar{\bar{v}}_j)/var(\bar{\bar{x}}_j)$. When μ is small, positive values of θ_1 compress the variance of x and compress the covariance of x and v, but less, such that $var(\bar{\bar{x}}_j) - var(\bar{x}_j) > cov(\bar{x}_j, \bar{\bar{v}}_j) - cov(\bar{\bar{x}}_j, \bar{\bar{v}}_j)$, which means that $\hat{\delta}_{agg} > \hat{\delta}'_{agg}$. When μ is large, positive values of θ_1 expand variance and $\hat{\delta}_{agg} < \hat{\delta}'_{agg}$. Similar logic applies to the case of $\theta_1 < 0$. Figure A3 in the online appendix demonstrates the compression and expansion of variance for different values of μ and θ_1.

Results for the individual model indicate no bias for any parameterization of μ and θ_1 because the exit poll sample comprises only voters and thus accounts for voter selection.

2.4 Combined effects of spatial omitted variables and voter selection

While Sections 2.2 and 2.3 examined sources of bias separately, we now combine the two sources, which is more likely to reflect actual data and applications. Figure 3 explores coefficient bias in the aggregate model and presents two panels: the left sets $\mu = 0.2$ and the right sets $\mu = 0.8$. The vertical axis of each plot measures coefficient bias and the horizontal axis is different values of $\theta_1 \in \{-0.2, 0, 0.2\}$. Three different values of ρ are plotted (-0.4, 0, 0.4).

Figure 3 illustrates that bias cannot be signed when both sources of bias are present. When working with real data, researchers could measure μ and gather additional data to approximate θ_1, and based on results in Section 2.3 could sign coefficient bias if voter selection was the only source of bias. Similarly, if a researcher had intuition about the sign of ρ, though unobserved by definition, then based on results of Section 2.2 they could sign coefficient bias if spatial omitted variable bias was the only source of bias. However, Figure 3 makes clear that just knowing the sign of each individual source of bias does not necessarily allow a researcher to sign the overall bias. The magnitudes of these two sources matter critically when they differ in sign, and the magnitudes are unobservable. Thus, in real applications, aggregate data will not only yield biased coefficients, but the bias may be of unknown sign or magnitude.

2.5 Exit poll participation selection

The exit poll relies on voluntary participation of voters. We now examine how selection into participation affects bias. The critical parameter is θ_2, the differential rate of exit poll participation for those with $x = 1$. For all iterations, we set $\rho = 0$ and $\theta_1 = 0$ to isolate this selection problem. We set $\omega \sim 80\%$ and $\mu = 0.2$, but these choices are inconsequential.

Figure 4 presents results. The vertical axis measures coefficient bias and the horizontal axis is different values of θ_2, ranging from -50% to 50%. The figure demonstrates zero bias in the exit poll coefficient, even in fairly extreme cases of one group being 50% more or less likely to take the survey. Intuitively, the regression compares mean approval across groups, and this quantity does not change if the size of one group gets bigger or smaller. The statistics that do suffer are estimates of voter participation by group and estimates of the referendum outcome. Applying sample weights based on measured versus actual vote outcomes can mitigate this. However, in the context of preference estimation, this imprecision is not as important as the regression coefficients.

In the online appendix, we additionally examine the cases of exit poll participation being correlated with approval and being correlated with an omitted variable that is also correlated with approval. Both of these variants also indicate zero bias in the exit poll coefficients.

2.6 Extensions of the Basic Data Generating Process

The DGP described above focuses on a single, binary characteristic. While helpful for clear exposition, we now discuss how the results generalize to more complex, realistic DGPs.[8] If a second demographic characteristic is added, the magnitude of bias for both aggregate and exit poll models is unchanged, but the variance in estimates from the aggregate model increases when correlation between the two independent variables is non-zero. We also investigate the case in

[8] These results are discussed in more detail in the online appendix, specifically Figures A7-13.

which the two independent variables have not just an additive effect on the dependent variable, but a multiplicative effect too. When a spatial omitted variable is present, the average bias of the main effect is similar to Figure 1 for both models, but the variance of the aggregate model grows. In terms of the interaction effect, the exit poll model is unbiased, but the aggregate model struggles with attenuation bias and large variance. Shifting to selection still in the context of multiplicative effects in the DGP, the exit poll model estimates are unbiased for both the main effect and the interaction effect. In contrast, the aggregate model estimates of the main effect exhibit similar bias as in the simple model, but the bias grows and can flip signs depending on the correlation between the two independent variables. The aggregate model estimates of the interaction effect suffer attenuation bias and increased variance.

In addition, we investigate different formulations of a single demographic characteristic. First, we modify the DGP to include a multinomial demographic characteristic instead of binary. We find similar results that bias in the aggregate model depends on distance from 50% population representation and the differential likelihood of voting. Second, we modify the DGP to include a continuous variable. The results indicate that any kind of voter selection leads to attenuation bias in the aggregate model. While this pattern is different than what is observed with discrete variables, it is consistent with the understanding of differences in variance between the total population and the voting population driving bias (see Section 2.3).

Summarizing the simulation results, in all cases we consider, the exit poll always performs at least as well as the aggregate model in terms of estimating unbiased voter preferences, and outperforms the aggregate model in most cases, including those that have realistic parameterizations.

3 CASE STUDY DATA

The case study focuses on a 2016 Rhode Island statewide bond referendum called the Green Economy Bonds (GEB). If approved, GEB would authorize the state to raise $35 million through bond sales with proceeds to be allocated for a suite of environmental priorities, including land conservation, brownfield remediation, stormwater pollution prevention, and bike paths. Bond repayment is financed through general revenue, which is primarily state income taxes and sales taxes. GEB passed with 67.6% approval, and 58.0% of the voting age population participated in the election. The goal of the empirical analysis is to estimate relative voter preferences of various partisan and socioeconomic groups for GEB.

3.1 Aggregate Data

From the Rhode Island Secretary of State, we obtained official vote tallies for GEB and the presidential race by precinct, of which there are 416. We also downloaded socioeconomic data at the Census block group level from the American Community Survey 2013-2017. From rigis.org, we downloaded a 2016 precinct shapefile, which we then overlaid with the block group shapefile to calculate area weights, which were then used to calculate the approximate socioeconomic mix for each precinct. This construction replicates the method of other research using aggregate data (e.g., Kotchen and Powers 2006, Banzhaf et al. 2010), though many papers use units of analysis larger than precincts.

We estimate the following descriptive regression model, which mirrors prior research:

$$\bar{\bar{v}}_j = \bar{\bar{\boldsymbol{P}}}_{\boldsymbol{j}}\boldsymbol{\delta_1} + \bar{\boldsymbol{X}}_{\boldsymbol{j}}\boldsymbol{\delta_2} + \boldsymbol{Z_j}\boldsymbol{\delta_3} + \varepsilon_j \qquad (1)$$

$\bar{\bar{v}}_j$ is the proportion of voters voting yes on GEB in precinct j, and $\bar{\bar{\boldsymbol{P}}}_{\boldsymbol{j}}$ is a vector of the proportions of voters voting for presidential contenders. Consistent with notation in Section 2, we use the double bar on these two variables to indicate a mean of the voting population. $\bar{\boldsymbol{X}}_{\boldsymbol{j}}$ is a vector of socioeconomic characteristics, each defined as the proportion of the population that is a given age, education level, gender, etc. A single bar is used here to denote that the average includes non-voters. The coefficients of interest are $\widehat{\boldsymbol{\delta}}_{\boldsymbol{1}}$ and $\widehat{\boldsymbol{\delta}}_{\boldsymbol{2}}$, which give the relative propensities of various partisan and socioeconomic groups to vote in favor of GEB. $\boldsymbol{Z_j}$ is a vector of location characteristics that may influence voting on GEB. Specifically, we include population density, 2016 residential property tax rate, average house sales price, acres preserved by state funds within 2km of precinct, precinct centroid distance to an existing bike path, and the number of remediated brownfields. Equation (1) is estimated by weighted least squares, with precincts weighted by the total number of votes for GEB officially recorded.

3.2 Exit Poll

We designed an exit poll survey to elicit votes for GEB and president and several socioeconomic characteristics (age, gender, race/ethnicity, income, education, and homeowner status).[9] All questions were multiple choice or yes/no, and age, income, and education questions intentionally used ranges that matched ranges of ACS variables to enable comparison. The questions were straightforward and phrased in plain language that followed the format used by national presidential exit polls, such as "In the Presidential election, who did you just vote for?". The complete survey instrument is provided in the appendix.

To implement the exit poll, we recruited 80 undergraduate and graduate student volunteers and placed them at 37 polling locations around the state for full day shifts. We chose sample poll locations to be representative of the state's socioeconomic characteristics, partisanship, and geographic scope. The online appendix provides additional details on the selection of sample precincts. While voters could vote early or by mail, there were significant hurdles to these options in Rhode Island in 2016, and only 8.5% of ballots cast were done in these ways.

Pollsters approached voters about survey participation as they were leaving the poll. The survey was completely anonymous and self-administered, which should mitigate social desirability bias. Pollsters reported an estimated 50% response rate. On Election Day, a total of 2,033 surveys were completed, which is the sample we use for analysis.[10]

We estimate the following model, which is the individual analogue of Equation (1):

$$v_{ij} = \boldsymbol{P_{ij}\delta_1} + \boldsymbol{X_{ij}\delta_2} + \boldsymbol{Z_j\delta_3} + \varepsilon_{ij} \qquad (2)$$

v_{ij} is binary and equals one if voter *i* in precinct *j* voted to approve GEB, $\boldsymbol{P_{ij}}$ is a vector of binary variables for presidential vote, $\boldsymbol{X_{ij}}$ is a vector of socioeconomic characteristics. As in Equation (1), $\widehat{\boldsymbol{\delta}}_{\boldsymbol{1}}$ and $\widehat{\boldsymbol{\delta}}_{\boldsymbol{2}}$ are the coefficients of interest. We estimate Eq. (2) using weighted least squares, with individual sampling weights determined by a person's votes for GEB and president (described below).

[9] An alternative source of individual data is a phone or internet survey, which is used for predictive election polling (e.g., CNN, Pew) and by some academic research. However, if the research question is explicitly about voters, exit polls have an advantage because the sample is necessarily all only voters. With other methods, it is necessary to generate a likely voter model, which is often inaccurate (Rogers and Aida 2014). Further, which voters respond to phone or internet surveys can change over time leading to inaccurate conclusions (Gelman et al. 2016).

[10] These data were previously used to investigate preferences for smart growth (dual approval of environmental preservation and affordable housing) (Pearson-Merkowitz and Lang 2020).

3.3 Summary Statistics

Table 1 provides means of voting and socioeconomic variables for both data sets. Column 1 presents means for the aggregate precinct data, which comprises the entire state of Rhode Island. The state leans liberal: GEB passed handily with 67.6% approval, and Hillary Clinton received 54.1% to Donald Trump's 39.2%. About two-thirds of residents are homeowners, just over half are female, and 18.8% are Hispanic or Black. The age distribution is fairly even. The income distribution shows a large mass of high earners, with 33.3% of households having an income above $100,000. In terms of education, just over one-third of the adult population is a high school graduate or dropout, 26.4% have some college, 21.5% graduate college, and 14.9% have a graduate degree. Column 2 gives means for just the 37 precincts that we sampled for the exit poll. These precincts mirror the state as a whole well, which was our intention in choosing sampling locations.

Our exit poll sample voted disproportionately for Clinton and for GEB relative to their precincts, with 63.3% for Clinton and 81.7% approval of GEB. Correlation between partisanship and exit poll participation is consistent with prior research (Best and Krueger 2012).[11] Given the results of Section 2.5 and the additional analysis in the online appendix, this selection is not a concern for estimating unbiased preferences. However, it does mean that demographic averages of the exit poll sample may not accurately reflect the averages of all voters in those precincts. To mitigate this inaccuracy, we calculate individual sampling weights that correct for the differential probability in exit poll participation. The weights equate voting percentages for the sample with the official results for the sampled precincts. Effectively, this gives a weight of less than one for those who voted for GEB or Hillary Clinton and a weight greater than one for those who voted against GEB or for Donald Trump, but also calculates a weight for every GEB-presidential vote combination. Given the correlation between demographics and vote choices, the weighted means provide a better estimate of the voting population characteristics.

[11] One concern specific to this election is that Trump voters may not honestly report their vote on our exit poll. The surprise outcome of the contest led many to believe that the pre-election polls were inaccurate due to "shy" Trump voters. Kennedy et al. (2018) examined many pre-election and post-election polls to assess this hypothesis. They argue that the pre-election polls relatively accurately predicted the popular vote. Forecasts "failed" in a handful of close states, but outcomes were within the margin of error. Instead, they find evidence of many late-deciding voters swinging for Trump, and post-election polls reflect this and the outcome accurately. Due to this conclusion and our use of an anonymous, self-administered survey, we are not too concerned about measurement error on the prudential vote. Empirically, while our survey respondents disproportionately voted for Clinton, they also disproportionately approved GEB and by similar margins, which reflects differences in willingness to participate in exit polls (Best and Krueger 2012) and not inaccurate responses.

Column 3 of Table 1 presents weighted means for the exit poll sample, and Column 4 presents differences in means between the exit poll and the population estimates of sampled precincts (Column 2). Differences in mean for the various votes cast are effectively set to zero through the application of weights, but even after weighting there are large differences in many of the socioeconomic variables. Individuals in households earning less than $30K are 8.8 percentage points less likely to be present in the exit poll sample than the population at large, whereas those coming from households earning $50-74K are 6.9 percentage points more likely to be present. The largest disparities in propensity to vote is seen by education level. Those with a high school degree or less are 21.8 percentage points less likely to be present in the exit poll sample, while those with a college or graduate degree are 9.2 percentage points more likely to be present.

The large disparities observed in Column 4 are a function of the selection process of who votes and the selection process of who responds to our survey. Leighley and Nagler (2013) document that the propensity to vote increases with age, income, and education. Because Column 4 patterns are consistent with these findings and we have applied weights to correct for uneven exit poll participation, we argue that Column 3 fairly accurately represents the voting population. Hence the large disparities observed in Column 4 suggest that the aggregate precinct data does not accurately measure the voting population. We can return to the Monte Carlo simulation results to assess coefficient bias in the aggregate model from mismeasurement alone. The Monte Carlo results indicate that the further the number in Column 4 is from zero and the further the number is in Column 2 from 50, then the larger the bias is, and the bias will be attenuation or amplification based on the sign in Column 4 and whether Column 2 is greater than or less than 50. For example, we would expect the aggregate model coefficient on household income less than $30K and education level high school or less to be attenuated because both groups comprise less than half the population and are less likely to vote. On the other hand, the aggregate model coefficients on household income $50-74K, education level college graduate and education level graduate degree are expected to be amplified because these groups comprise less than half the population and are more likely vote.

However, the bias predictions based on ignoring voter selection may not hold because the second source of bias, omitted spatial variables, may disrupt those patterns, as shown in Figure 3. While we of course cannot present statistics on possible omitted variables, one key element determining the consistency of bias in the aggregate model and the amount of bias in the individual exit poll model is the amount of within-precinct variation. Column 5 presents this statistic for each variable derived from the exit poll.[12] For all age and education groups, within-precinct variation is above 95%. For middle income groups, it is over 97%, while it is 92% for less than $30K and 90% for over $100K. Even for presidential votes within-precinct variation is large at 90%, which is surprising given widespread evidence of sorting and polarization (e.g., Card et al. 2008, Lang and Pearson-Merkowitz 2015). We see the most sorting for Hispanic and Black, but still 77% of variation is within precinct. The second lowest is for homeowners at 86% within.

Given the high degree of within-precinct variation, we expect most variables in the individual exit poll model to be relatively immune from omitted spatial variable bias. Further, these numbers establish how little of the total variation is used for estimation in the aggregate model.

4 Case Study Results

Table 2 presents voter preference regression results for both datasets. For both columns, the coefficients are interpreted the same: the percentage point change in likelihood of voting for GEB resulting from a one percentage point increase in that variable. For age, income, and education, which all have multiple categories, the omitted category is chosen as the one with the smallest difference between population mean and exit poll mean in order to minimize the bias affecting the omitted group.[13] Precinct characteristics ($\boldsymbol{Z_j}$) are included in both models, but not displayed.

Coefficients are considerably different across datasets. Consider first the coefficients on presidential voting. The coefficients on Voted for Hillary Clinton are both large, positive, and precisely estimated, implying that Democrats prefer environmental spending more than Republicans. However, the aggregate model coefficient is 32% larger than the individual model,

---

[12] Ideally, this would be measured at the population level. However, Monte Carlo evidence suggests that exit poll estimates of within-precinct variation are good approximation of population within-precinct variation, even in the presence of differential voting participation rates by groups.

[13] If bias resulting from ignoring voter selection (mismeasurement of voter characteristics) affects the omitted group, then all estimated coefficients that are relative to that group will also be biased even if there is no mismeasurement for the other groups. For income, we use over $100K as the omitted group even though the difference in means is slightly larger than $30-49K because proportion of population with income over $100K is over twice as large as $30-49K and standard errors are smaller as a result.

and the point estimate from the exit poll would be rejected with over 99% confidence by the aggregate model. Coefficients on votes for third-party candidates are even more disparate across models. The aggregate model finds that Gary Johnson supporters are 46 percentage points more likely to vote for GEB than Trump voters, a similar margin as Clinton voters. However, the individual exit poll finds that Johnson voters are equally likely to vote for GEB as Trump voters, which is more intuitive given ideological stances of candidates and libertarians beliefs in small government. The aggregate model coefficient on Voted for Jill Stein is over 100 percentage points larger than the individual model and is implausible when interpreted as an individual preference: one vote for Jill Stein leads to 1.2 votes for GEB. Votes for alternative candidates are especially susceptible to omitted spatial variables because of how little between-precinct variation exists (less than 2% of total variation).

Socioeconomic characteristics yield different coefficients across models too. The coefficients on homeowner are both negative and highly significant, but the individual model coefficient is nearly twice as large in magnitude and the point estimate would be rejected with 99% confidence by the aggregate model. The aggregate model coefficient on Age 65 or over is -0.079 and is statistically significant, but the individual model yields an insignificant coefficient of -0.023. The aggregate model coefficients on income less than $30K, $30-49K, and $50-74K are all negative and statistically significant, whereas no such pattern emerges in the individual model. The models also paint a very different picture in terms of preferences of people with different education levels. The aggregate model suggests that people with a high school degree or less and those with a college degree are more likely to vote for GEB than those with some college, but the coefficient on graduate degree is statistically zero. In complete contrast, the individual model indicates that those with a graduate degree are more likely to approve GEB, and the coefficients on high school degree or less and college degree are statistical zeros.

There are no other papers that execute the same type of analysis with both aggregate data and individual exit poll data, so a comparison of these results to prior literature is difficult. The most similar is Holian and Kahn (2015) that analyzes both aggregate voting data and an individual telephone survey, with both focused on two statewide referendums in California. While the authors interpret the results as corroborating each other, there are strong differences across datasets and some of the same trends appear as in this paper. For example, Holian and Kahn find that both data sets yield coefficients on presidential vote that are the same sign and both highly statistically significant, but the coefficient on voting for Bush in the aggregate model is 200% larger than the corresponding coefficient in the individual model for one referendum, and 50% larger for the second referendum. Further, almost all coefficients (e.g., income, race and age) change sign or significance or have a large change in magnitude across models. In a different context, Thalmann (2004) and Bornstein and Lanz (2008) analyze voter preferences on the same three Swiss referendums, but do so with different data. Thalmann uses an individual survey and Bornstein and Lanz use official precinct votes matched with Census data. The two datasets do not agree about estimated preferences for different age groups and language groups. A third point of comparison is Wu and Cutter (2011), who also examine voting on multiple California referendums. They estimate models at different levels of aggregation (block group, tract, county), and find large differences in coefficients across models, which they interpret as aggregation bias. Vossler and Kerkvliet (2003) actually find similar results using individual survey data and aggregate voting data, but they make an important methodological choice. Instead of relying on Census data in the aggregate model, they use the individual survey and additional data from individual voters to construct the precinct-level demographic means, which essentially eliminates the voter selection issue.

Table 2 establishes differences in the preference estimates of the two datasets. The online appendix presents several additional specifications, including using a logit model instead of linear probability model, and results are robust throughout. The following two sections seek to further establish that omitted spatial variables and voter selection are sources of bias in the aggregate model.

4.1 Assessing the role of spatial omitted variable bias

We cannot directly test for the presence of spatial omitted variables by definition. Instead, we indirectly test for their presence by examining how coefficients change as various spatial variables are included. If coefficients do not change as spatial variables are included, then we judge the role of spatial omitted variable bias to be minimal.

Table 3 Columns 1-4 presents four different specifications using the aggregate precinct data. Column 1 includes only presidential votes and socioeconomic characteristics as independent variables, Column 2 adds precinct characteristics (this is the main specification from Table 2), Column 3 adds town fixed effects, and Column 4 adds quadratic functions of latitude and longitude of precinct centroids.[14] Looking across columns at coefficients on presidential votes, there are small increases in the coefficients on Hillary Clinton and Gary Johnson as spatial covariates are added. The coefficients on Jill Stein and Other both increase substantially when town fixed effects are included. The coefficients on Homeowner and Age 65 or over decline in magnitude somewhat, but maintain their statistical significance as spatial variables are added. The largest changes are observed in the household income and education level variables. Coefficients on income less than $30K, income $30-49K, income $50-74K, high school education or less, and college degree all sharply decline in magnitude and lose significance when town fixed effects are included.

These large changes observed in several variables suggest that the aggregate model independent variables are correlated with unobserved spatial variables. Particularly when town fixed effects are included, we see large changes in coefficient magnitudes and statistical significance. However, we see mixed evidence of whether adding spatial covariates moves aggregate coefficients closer to the individual coefficients. Income and education coefficients that are statistically significant move towards their individual model counterparts, but they also move towards zero, so the shift could merely reflect a reduction in the variance available for estimation. Further, the coefficient on graduate degree does not move at all, and the homeowner and presidential vote coefficients move further away. Thus, adding spatial controls is not a panacea and aggregate coefficients remain biased even with that addition.[15]

Table 4 presents results from the individual model in the same vein as Table 3. Four specifications are given, with each column beyond the first adding spatial control variables. Column 1 includes only presidential votes and socioeconomic characteristics as independent variables. Column 2, which replicates Table 2 results, adds precinct characteristics. Column 3 adds town fixed effects, and Column 4 adds precinct fixed effects. Hence, Column 4 uses entirely within-precinct variation to estimate coefficients. In contrast to Table 3, the results show little change in coefficient magnitudes across columns and only one instance of lost statistical

[14] Rhode Island has 39 towns, with an average of 10.7 precincts per town, and the median town has 7 precincts.
[15] If spatial omitted variables are time-invariant, then using multiple vote outcomes for similar referendums and spatial fixed effects could alleviate bias. Some research has used this approach, though not always with the stated intent to control for unobservables. Brunner et al. (2011) and Altonji et al. (2016) essentially construct panel data with repeated vote approval observations by Census tract and include tract fixed effects in their regressions. Anderson et al. (2019) use changes in vote shares across two referendums to examine how changes in referendum design affects approval.

significance. These results are intuitive because of the large proportion of variation that is within-precinct, and hence not susceptible to spatial omitted variable bias. Additionally, the results support the Monte Carlo findings that for high levels of within-precinct variation, the individual exit poll model is unlikely to be biased by spatial omitted variables.

4.2 Assessing the role of voter selection

The aggregate model ignores the selection process of who decides to vote and in doing so mismeasures voter characteristics. In this section, we first present evidence of how mismeasurement of voter characteristics can cause biased coefficients by inducing mismeasurement in the exit poll data. Second, we test whether a simple adjustment of population characteristics in the aggregate model can reduce bias.

Using the individual exit poll data, we develop a procedure that induces systematic measurement error so that the demographic means of the exit poll sample match the Census population. For individuals with characteristics that are over- or under-represented among voters relative to the total population, we randomly reassign those individuals' characteristics until the sample mean equals the population mean. For example, relative to the general population, voters are more likely to have a college degree or graduate degree and less likely to have a high school degree or less. We randomly and iteratively assign sampled voters that have a college degree or graduate degree to instead have only a high school education until the proportions among voters and general population are the same across all education levels. We do the same reassignment for age, income, homeowner, female and Hispanic or Black. After reassignment, we estimate the main individual exit poll regression model from Table 2. We repeat the process of randomized reassignment and estimation 100 times in order to estimate a distribution of the coefficients with induced measurement error. By doing this exercise with the individual data, we are essentially purging the issue of omitted spatial variable bias and isolating bias due to ignoring voter selection.

Figure 5 plots the distribution of the coefficients with induced mismeasurement for a subset of variables, as well as the point estimates of the individual exit poll and aggregate precinct models from Table 2 for reference.[16] In four of the six variables displayed, the distribution moves towards the aggregate coefficient and away from individual coefficient. The

[16] Table A4 in the online appendix provides means and distribution statistics for all coefficients in the model.

largest shift is seen for high school or less education, which has a mean about halfway between the individual and aggregate coefficients. It is intuitive that we see the largest shift with this variable as it had the largest difference between exit poll sample and general population. Even the coefficient on voting for Hillary Clinton shifts towards the aggregate coefficient, and there was no reassignment of presidential votes, so this shift comes entirely from reassignment of other variables. This figure also illustrates how the systematic mismeasurement of the voting population differs from classical measurement error. Adding random noise to an independent variable causes attenuation bias in the coefficients. However, in two cases (Clinton, high school or less), the distribution of coefficients with induced measurement error shifts to be larger in magnitude.[17] In sum, this figure shows that ignoring voter selection indeed accounts for some of the gap between the preference estimates of the exit poll and aggregate model.

We now attempt a solution to the mismeasurement problem of the aggregate model by adjusting demographics to reflect the selection process. We adjust all Census demographic variables such that the mean matches the exit poll means. For each characteristic, we calculate the ratio of exit poll mean to sampled precinct population mean (Column 3 divided by Column 2 in Table 1) and multiply the precinct values of the characteristic by that ratio. For example, for household income $30K or less, the factor is 12.2/21.0 = 0.581, so a precinct with 30% of the population with high school or less education is adjusted to have an estimated 17.4% (=30%*0.581) high school or less education among the voting population. All precincts are adjusted using the same factors.

Column 5 of Table 3 presents results after the demographic adjustment. The specification is identical to Column 2 of Table 3, the only difference is the adjustment, so Column 2 is the relevant comparison. The adjustment has little impact on coefficients on presidential votes. However, the adjustment does impact coefficient magnitudes on socioeconomic characteristics, though coefficient sign never changes. The coefficients on homeowner, household income $50-74K, college graduate, and age 65 or over become slightly smaller in magnitude. Some changes are larger though; the coefficient on household income $30K or less nearly doubles in magnitude, and the coefficient on education high school or less nearly triples. While covariate adjustment clearly affects coefficients, more often than not the coefficients move further from

[17] In other cases it is not possible to distinguish shifts in the distribution from attenuation because the aggregate coefficient is either smaller in magnitude or a different sign than the individual coefficient.

the individual exit poll coefficients, rather than closer. Even with adjustments to demographics, the aggregate model still relies solely on between-precinct variation, which we argue incapacitates the aggregate model from generating unbiased estimates.[18]

4.3 Impact of bias on valuation and distributional impact estimates

In the previous several pages, we have established that aggregate model coefficients are biased and the reasons for that bias. This result by itself indicates that regression estimates using aggregate data give an incorrect assessment of various groups' propensity to vote for or against a given referendum. However, in addition to estimating willingness to vote, referendum outcomes can be used to estimate willingness to pay (WTP) for the proposed amenities or services (e.g., Vossler and Kerkvliet 2003, Burkhardt and Chan 2017). These WTP estimates can then be used in cost-benefit analysis to assess the efficiency and distributional implications of a policy, or for predicting the distribution of benefits of future spending. In this section we explore how the data type can influence WTP estimates.

Following referendum-style contingent valuation methods (Hanemann 1984, 1989), cost of the referendum must be included as an independent variable in the regression in order to derive WTP (Burkhardt and Chan 2017). The GEB will be paid for through general revenue, which is primarily income and sales taxes. We approximate household cost of GEB based on Rhode Island specific incidence of state income and sales taxes by income group from the Institute on Taxation and Economic Policy [ITEP] (2018).[19] We calculate that the cost of GEB for a household with the median income of $61,000 is about $82.03, and cost increases on average $12.07 for every $10,000 of additional income. Because estimated cost is highly collinear with income and homeowner status, these variables are excluded from the regression. Due to these necessary changes in the voter choice model, our WTP estimates are purely meant to illustrate possible disparities in estimates across data sources.[20]

---

[18] In the online appendix, we estimate a voter participation model using aggregate data, presented in Table A5. In addition to the strategy of adjusting precinct demographic levels shown in Table 3, we additionally adjust means using estimated propensities for different groups to vote. Neither this approach, nor including residuals from the participation model, nor including participation levels has a substantial impact on the aggregate model voter preference estimates. This reinforces our belief that a bias-correction solution does not exist for the aggregate model.

[19] While state income tax rates are progressive, the incidence of sales taxes is regressive. ITEP (2018) calculates on net that state taxes are regressive, with the bottom quintile of the income distribution paying 6.5% of income in taxes and the top quintile paying about 5.2%.

[20] Referendum-style contingent valuation surveys randomize hypothetical cost, and thus cost is orthogonal to income, which obviates this issue.

Table 5 presents estimated WTP for the aggregate and individual models, as well as the deviation from average WTP for various groups, which also represents the distribution of benefits from GEB passage.[21] The estimated WTP using aggregate precincts is $848, which is 2.6 times larger than the estimate using the individual exit poll ($323). The standard error in the aggregate model is large however, and the aggregate model does not reject the point estimate from the individual model. The individual model is more precisely estimated. In terms of the estimated distribution of benefits, often the models agree about which groups benefit more, but the magnitude can be different. Both models agree that Democrats have a larger WTP, but the estimated spread between Democrats and Republicans is about 67% larger for the aggregate model than the individual model, suggesting larger partisan divides in valuation than actually exist. Further, the aggregate model indicates that Republicans are actually harmed by GEB environmental improvements (WTP=-$432), whereas the individual model indicates that Republicans still value the improvements, just much less than average (WTP=$38). For age groups, the models agree on sign, but the aggregate model indicates a larger negative deviation for the old than the positive deviation for the young, but the individual model estimates that the deviation for the young is greater than that for the old. The most meaningful difference in models is for estimated benefit distributions between education groups. While the aggregate model suggests those with high school education or less have a higher WTP than those with a graduate degree, the individual model estimates the opposite and indicates a larger spread in WTP between these groups. The models also disagree about the distribution of benefits between racial and ethnic groups, with the aggregate model finding that Hispanics and Blacks benefit less than whites, whereas the individual model finds the opposite.

Returning to the initial motivation for estimating WTP and distributional impacts, Table 5 illustrates that using aggregate data could lead to incorrect assessment of efficiency and equity of a proposed or actual referendum, and thus an inefficient provision of public goods. For example, if a government is interested in creating environmental policy to address or at least consider racial injustice, the aggregate analysis may steer them away from spending similar to

[21] Table A6 in the online appendix provides regression results used to estimate WTP. Linear models are used similar to Table 2. In order to calculate WTP, we modify the Hanemann's (1984, 1989) logit formula for use with LPM. Average WTP is calculated as $(\hat{\alpha} + \overline{\boldsymbol{X}}\widehat{\boldsymbol{\delta}} - 0.5)/(-\hat{\gamma}_{cost})$, where $\hat{\alpha}$ and $\widehat{\boldsymbol{\delta}}$ come from the regression, $\overline{\boldsymbol{X}}$ is the mean of all independent variables (other than cost) included in the regression, and $\hat{\gamma}_{cost}$ is the estimated coefficient on cost. Distributional impacts are calculated by conditioning on specific values to $\boldsymbol{X}$ instead of sample means. WTP estimates derived from logit models and the standard Hanemann equation are near identical.

GEB priorities. To be clear, it is not always the case that the aggregate model will overestimate WTP or underestimate differential benefits for minorities, these result are specific to this context. But given the biases that the aggregate model possesses, it is quite likely that WTP and distributional conclusions will be wrong in other contexts.

5 CONCLUSIONS

Understanding voter preferences and valuation of public goods and the distributional consequences of referendums are critical research areas. Many papers have developed revealed preference models that use publicly available aggregated votes matched with Census demographic data to address this research need. In this paper, we explore two potential sources of bias of the aggregate data approach, omitted spatial variables and ignoring voter selection, and then examine these sources through both a Monte Carlo simulation and a case study involving an actual statewide referendum.

In the simulation, we are able to cleanly confirm that both potential sources of bias do indeed lead to biased preference estimates when aggregate data are used. With only one source of bias present, the sign of bias is known. However, with both sources present, which is the case in real-world applications, the sign of bias in unknown. In addition to the aggregate data, we also model a hypothetical exit poll that gathers individual level information on a small subset of voters. The exit poll only samples from the voting population, so avoids one source of bias, and can rely on within-precinct variation to mitigate bias from omitted spatial variables. Even in the presence of exit poll participation selection issues, the exit poll is always less biased than the aggregate model, and typically unbiased under realistic parameters.

In the case study of a Rhode Island statewide environmental bond referendum, we mimic prior research and match aggregated, official votes to Census data at the precinct level and additionally undertake an extensive, statewide exit poll that is used for comparison. Results demonstrate large differences between preference estimates based on aggregate data and those based on exit polls, which we interpret as biased aggregate coefficients. While we cannot cleanly separate sources of bias, we perform additional tests that indicate both spatial omitted variables and ignoring voter selection lead to bias in the aggregate model. Lastly, we demonstrate that aggregate data produce an invalid assessment of valuation and distributional impacts.

The conclusion of this paper is that analysis of aggregate voting data matched with Census data is unlikely to yield unbiased voter preference estimates, and should be used only with this caveat. However, our results identify conditions when bias will likely be strongest, which can help future studies. For analysis of statewide referendums, states that are larger or more heterogeneous in terms of demographics and economy may have more unobserved determinants of approval, and hence aggregate data would be more susceptible to bias. California falls into this category and has been the focus of many studies due to the availability of data and the prevalence of referendums (e.g., Wu and Cutter 2011, Holian and Kahn 2015, Burkhardt and Chan 2017). Other research has compared votes on similar referendums across municipalities throughout the United States (e.g., Kotchen and Powers 2006, Banzhaf et al. 2010). One concern with this strategy is that local referendums are more likely than statewide referendums to be held in off-cycle (non-November) elections, which typically means even lower voter turnout.[22] This in turn suggests voter selection forces may be stronger and disparities between the voting population and general population may be larger, which would increase bias when using aggregate data. Thus, analyzing voter preferences on a single citywide or statewide referendum in a small homogenous city or state held during a presidential election is when aggregate data will have the least bias. However, these are the conditions of our case study in Rhode Island, and bias is still substantial.

While we are pessimistic about the prospects of standard aggregate voting analysis, this paper presents the exit poll as a viable alternative. We have demonstrated that this method can be operationalized to cover an area the size of most MSAs and do so for a low monetary cost. We hope that others apply this method in other states and other referendums to replicate our findings and address important questions of voter behavior and political economy. However, given the broader use of early and mail-in voting, greatly accelerated by COVID-19, exit polls may become less viable. Other individual-level surveys by telephone, mail, or internet, are still feasible and commonly used in valuation settings.

[22] According to the Land Vote Database, which is the data source for several studies on voter preferences for land conservation, only 22.3% of municipal referendums are held on presidential general election days, 20.2% are held on midterm election days, and 57.5% are held at other times. While about 55-60% of the voting age population (VAP) participate in presidential elections, midterm elections typically draw 10-15% less and only 25% or less of the VAP may participate in local elections (Hajnal and Lewis 2003). Thus, mismeasurement bias is likely worst for off-cycle municipal elections.

Future methods may be developed, perhaps using big data or machine learning, to mitigate bias in aggregate models. Ghitza and Gelman (2018) develop a methodology that combines telephone surveys and voter registration databases to predict presidential and congressional votes for individuals. While promising, this is not a method for aggregate data alone, the survey is the workhorse for prediction, which suggests there is a fundamental necessity for individual-level variation.

References

Altonji, M., Lang, C., & Puggioni, G. (2016). Can urban areas help sustain the preservation of open space? Evidence from statewide referenda. *Ecological Economics*, *130*, 82-91.

Anderson, S. T., Marinescu, I., & Shor, B. (2019). *Can Pigou at the Polls Stop Us Melting the Poles?* (No. w26146). National Bureau of Economic Research.

Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives*, *33*(1), 185-208.

Banzhaf, H. S., Oates, W. E., & Sanchirico, J. N. (2010). Success and design of local referenda for land conservation. *Journal of Policy Analysis and Management*, *29*(4), 769-798.

Best, Samuel J., and Brian S. Krueger. (2012). *Exit polls: Surveying the American electorate, 1972-2010*. CQ Press.

Bornstein, N., & Lanz, B. (2008). Voting on the environment: Price or ideology? Evidence from Swiss referendums. *Ecological Economics*, *67*(3), 430-440.

Brunner, E. J., & Ross, S. L. (2010). Is the median voter decisive? Evidence from referenda voting patterns. *Journal of Public Economics*, *94*(11-12), 898-910.

Brunner, E., Ross, S. L., & Washington, E. (2011). Economics and policy preferences: causal evidence of the impact of economic conditions on support for redistribution and other ballot proposals. *Review of Economics and Statistics*, *93*(3), 888-906.

Burkhardt, J., & Chan, N. W. (2017). The dollars and sense of ballot propositions: estimating willingness to pay for public goods using aggregate voting data. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 479-503.

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics*, *123*(1), 177-218.

Cho, W. K. Tam, & Gaines, B. J. (2004). The limits of ecological inference: the case of split-ticket voting. *American Journal of Political Science*, *48*(1), 152-171.

Cutler, D. M., Elmendorf, D. W., & Zeckhauser, R. (1999). Restraining the Leviathan: property tax limitation in Massachusetts. *Journal of Public Economics*, *71*(3), 313-334.

Deacon, R. T., & Schläpfer, F. (2010). The spatial range of public goods revealed through referendum voting. Environmental and resource economics, 47(3), 305-328.

Deacon, R., & Shapiro, P. (1975). Private preference for collective goods revealed through voting on referenda. *The American Economic Review*, 943-955.

DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, *84*(1), 143-181.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American sociological review*, 557-572.

Fischel, W. A. (1979). Determinants of voting on environmental quality: a study of a New Hampshire pulp mill referendum. *Journal of Environmental Economics and Management*, *6*(2), 107-118.

Gelman, A., Goel, S., Rivers, D., & Rothschild, D. (2016). The mythical swing voter. *Quarterly Journal of Political Science*, *11*(1), 103-130.

Gerber, A. S., & Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American political science review*, 653-663.

Gerber, E. R., & Lewis, J. B. (2004). Beyond the median: Voter preferences, district heterogeneity, and political representation. *Journal of Political Economy*, *112*(6), 1364-1383.

Ghitza, Y., & Gelman, A. (2018). Voter Registration Databases and MRP: Toward the Use of Large-Scale Databases in Public Opinion Research. *Political Analysis*, 1-25.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632-648.

Greenland, S., & Morgenstern, H. (1989). Ecological bias, confounding, and effect modification. *International journal of epidemiology*, *18*(1), 269-274.

Greiner, D. J., & Quinn, K. M. (2010). Exit polling and racial bloc voting: Combining individual-level and R x C ecological data. *The Annals of Applied Statistics*, 1774-1796.

Hajnal, Z. L., & Lewis, P. G. (2003). Municipal institutions and voter turnout in local elections. *Urban Affairs Review*, *38*(5), 645-668.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American journal of agricultural economics*, *66*(3), 332-341.

Hanemann, W. M. (1989). Welfare evaluations in contingent valuation experiments with discrete response data: reply. *American journal of agricultural economics*, *71*(4), 1057-1061.

Holian, M. J., & Kahn, M. E. (2015). Household demand for low carbon policies: Evidence from California. *Journal of the Association of Environmental and Resource Economists*, *2*(2), 205-234.

Institute on Taxation and Economic Policy [ITEP]. (2018). *Who Pays? A distributional analysis of the tax system in all 50 states, sixth edition*. Downloaded from https://itep.org/wp-content/uploads/whopays-ITEP-2018.pdf

Kahn, M. E., & Matsusaka, J. G. (1997). Demand for environmental goods: Evidence from voting patterns on California initiatives. *The Journal of Law and Economics*, *40*(1), 137-174.

Kennedy, C., Blumenthal, M., Clement, S., Clinton, J. D., Durand, C., Franklin, C., ... & Wlezien, C. (2018). An evaluation of the 2016 election polls in the United States. *Public Opinion Quarterly*, *82*(1), 1-33.

King, Gary. *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton University Press, 2013.

Kotchen, M. J., & Powers, S. M. (2006). Explaining the appearance and success of voter referenda for open-space conservation. *Journal of Environmental Economics and Management*, *52*(1), 373-390.

Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?. *Journal of Environmental Economics and Management*, *60*(3), 145-160.

Lang, C., & Pearson-Merkowitz, S. (2015). Partisan sorting in the United States, 1972–2012: new evidence from a dynamic analysis. *Political Geography*, *48*, 119-129.

Leighley, J. E., & Nagler, J. (2013). *Who votes now?: Demographics, issues, inequality, and turnout in the United States*. Princeton University Press.

Matsa, D. A., & Miller, A. R. (2018). *Who Votes for Medicaid Expansion? Lessons from Maine's 2017 Referendum* (No. w25109). National Bureau of Economic Research.

Matsusaka, J. G. (2005). Direct democracy works. *Journal of Economic perspectives*, *19*(2), 185-206.

McDonald, M. P., & Popkin, S. L. (2001). The myth of the vanishing voter. *American Political Science Review*, *95*(4), 963-974.

Pearson-Merkowitz, S., & Lang, C. (2020). Smart Growth at the Ballot Box: Understanding Voting on Affordable Housing and Land Management Referendums. *Urban Affairs Review*, *56*(6), 1848-1875.

Piantadosi, S., Byar, D. P., & Green, S. B. (1988). The ecological fallacy. *American journal of epidemiology*, *127*(5), 893-904.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, *15*(3), 351-357.

Rogers, T., & Aida, M. (2014). Vote self-prediction hardly predicts who will vote, and is (misleadingly) unbiased. *American Politics Research*, *42*(3), 503-528.

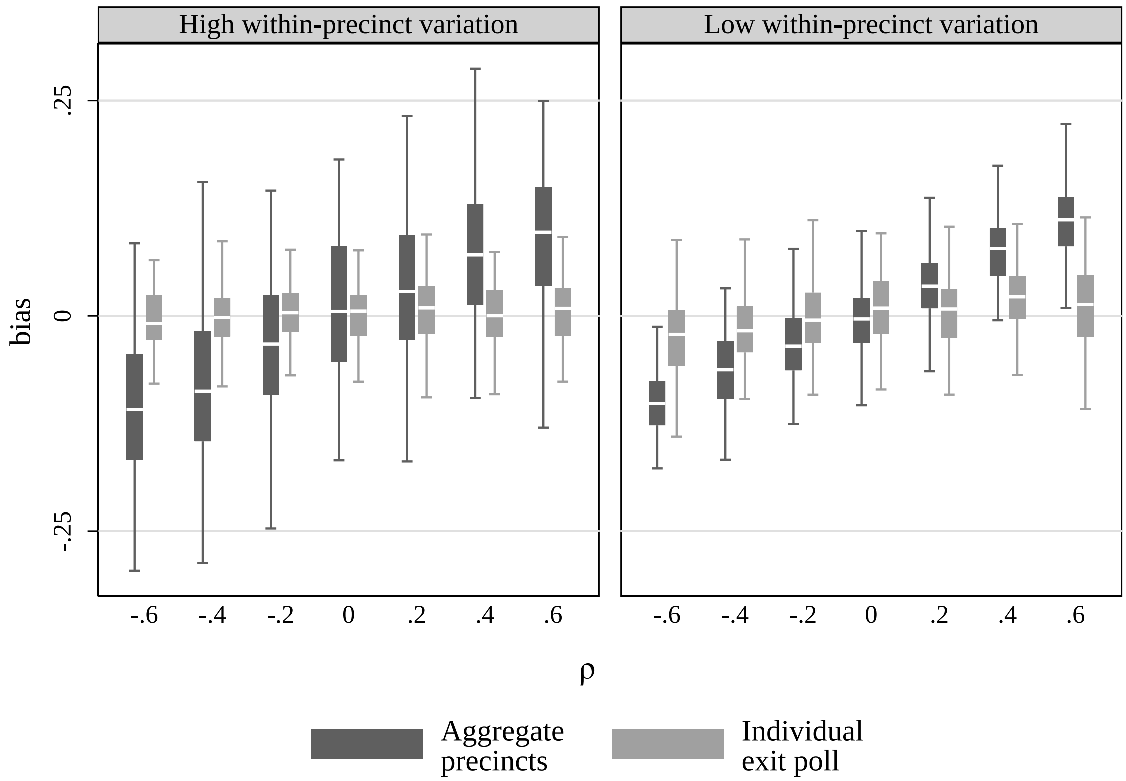
Thalmann, P. (2004). The public acceptance of green taxes: 2 million voters express their opinion. *Public Choice*, *119*(1-2), 179-217.

Vossler, C. A., & Kerkvliet, J. (2003). A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum. *Journal of Environmental Economics and Management*, *45*(3), 631-649.

Wu, X., & Cutter, B. (2011). Who votes for public environmental goods in California?: Evidence from a spatial analysis of voting for environmental ballot measures. *Ecological Economics*, *70*(3), 554-563.

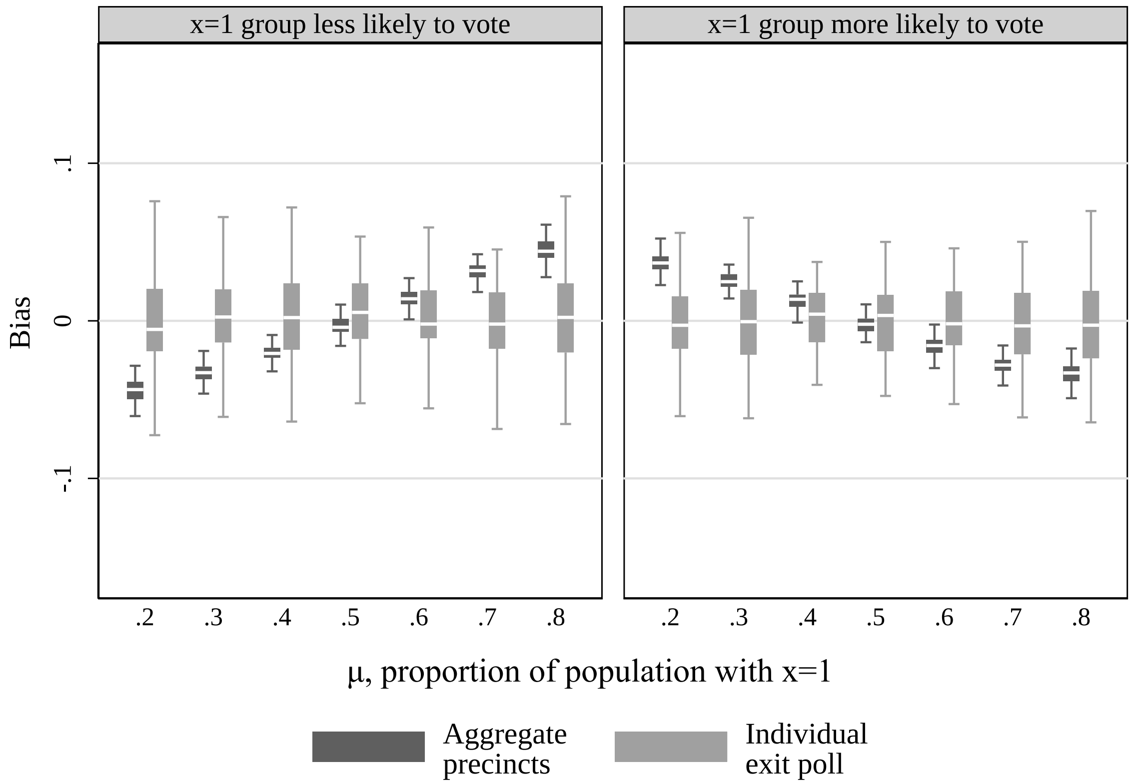
Figures and Tables

Figure 1: Effect of spatial omitted variable and within precinct variation on coefficient bias



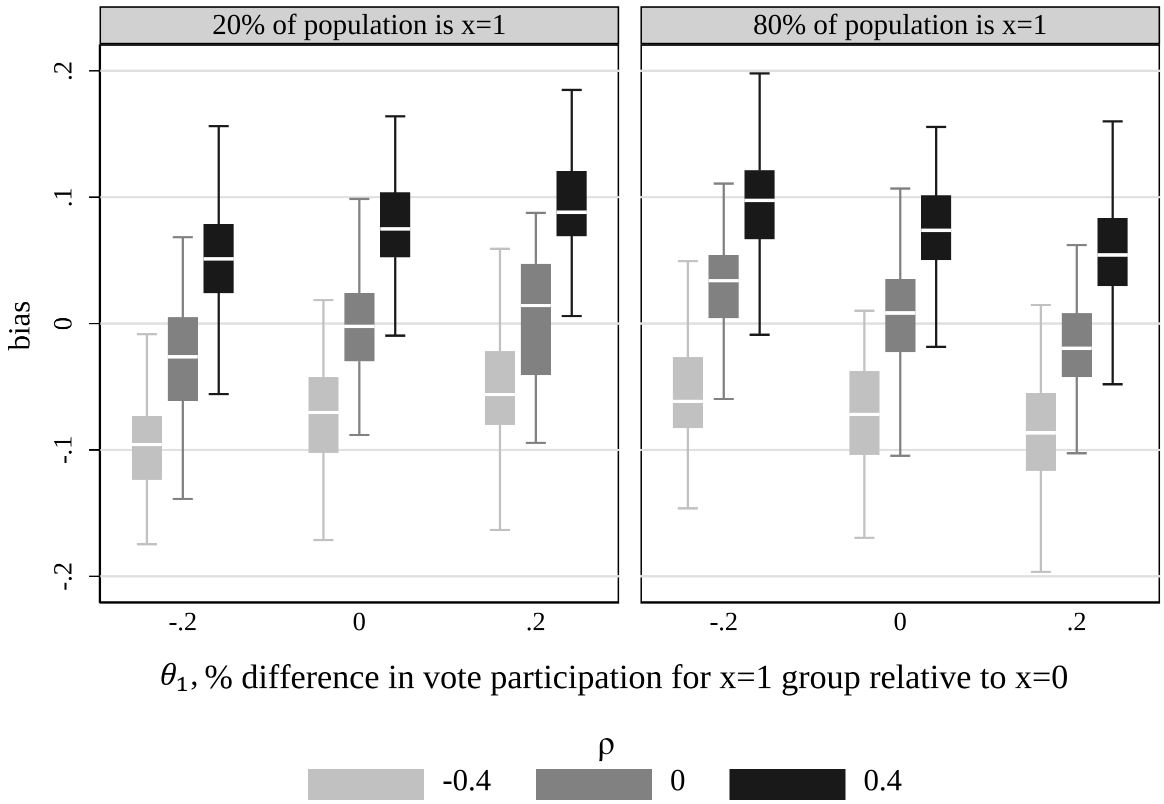
Notes: Results come from Monte Carlo simulation described in text. Vertical axis is the difference between the unbiased estimate and the estimate derived from either the aggregate data or exit poll data. Horizontal axis is different values of parameter ρ; negative (positive) values represent a negative (positive) correlation between a spatial omitted variable and the variable of interest, x. For the left plot, mean ω is 95.7% (sd=0.81), and for the right plot, mean ω is 79.9% (sd=1.21). Other parameters are set as follows: $N = 250{,}000$, $J = 500$, $N_j = 500$, $p = 0.1$, $\pi_1 = 0.5$, $\pi_2 = 0.05$, $\beta = 1$, $\sigma = 2$, $\mu = 0.2$, $\theta_1 = 0$ and $\theta_2 = 0$. Box plot boundaries indicate 25^{th}-75^{th} percentile with the middle line being the median and the whiskers are the 5^{th} and 95^{th} percentiles. The mean of the unbiased estimates is 0.176.

Figure 2: Effect of selection into voting on coefficient bias



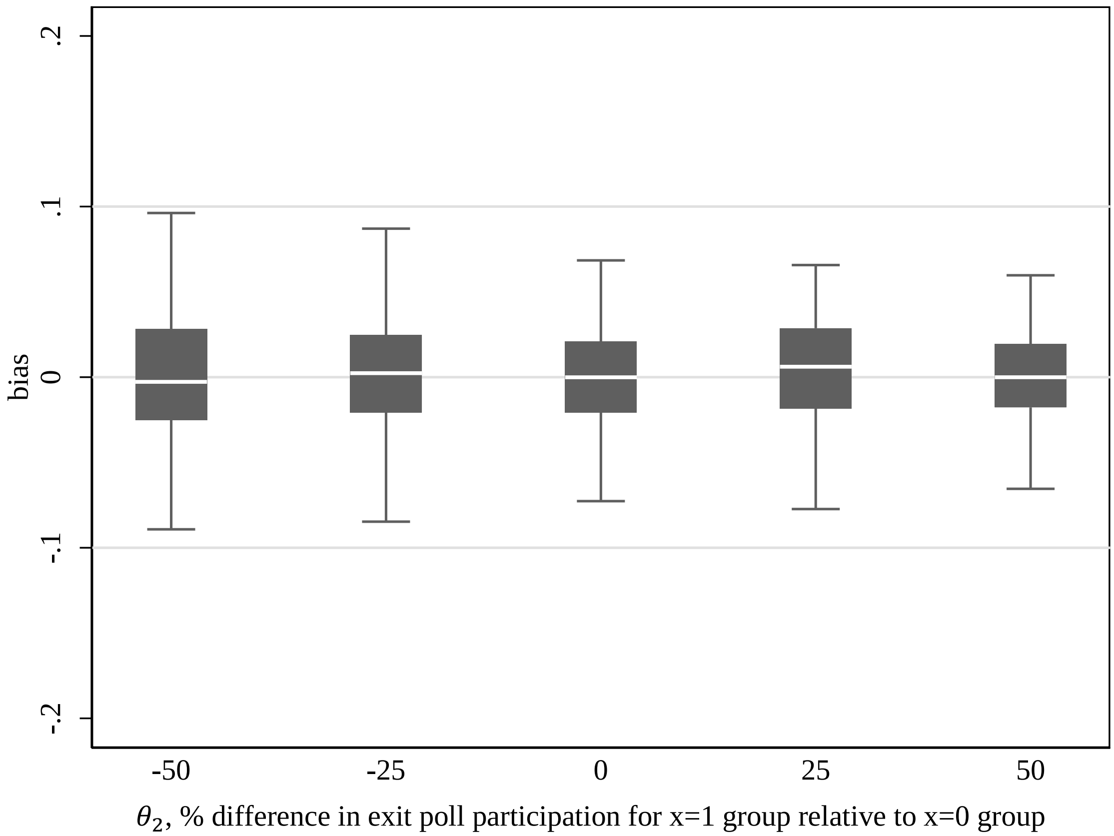
Notes: Results come from Monte Carlo simulation described in text. Vertical axis is the difference between the unbiased estimate and the estimate derived from either the aggregate data or exit poll data. The left plot sets $\theta_1 = -0.2$, and the right plot sets $\theta_1 = 0.2$. Box plot boundaries indicate 25th-75th percentile with the middle line being the median and the whiskers are the 5th and 95th percentiles. The mean of the unbiased estimates is 0.176. Other parameters are set as follows: $N = 250{,}000$, $J = 500$, $N_j = 500$, $p = 0.1$, $\pi_1 = 0.5$, $\pi_2 = 0.05$, $\theta_2 = 0$, $\beta = 1$, $\sigma = 2$, $\omega \sim 80\%$, $\rho = 0$, and $\tau_j = 0$.

Figure 3: Effect of selection into voting and spatial omitted variables on coefficient bias in aggregate model



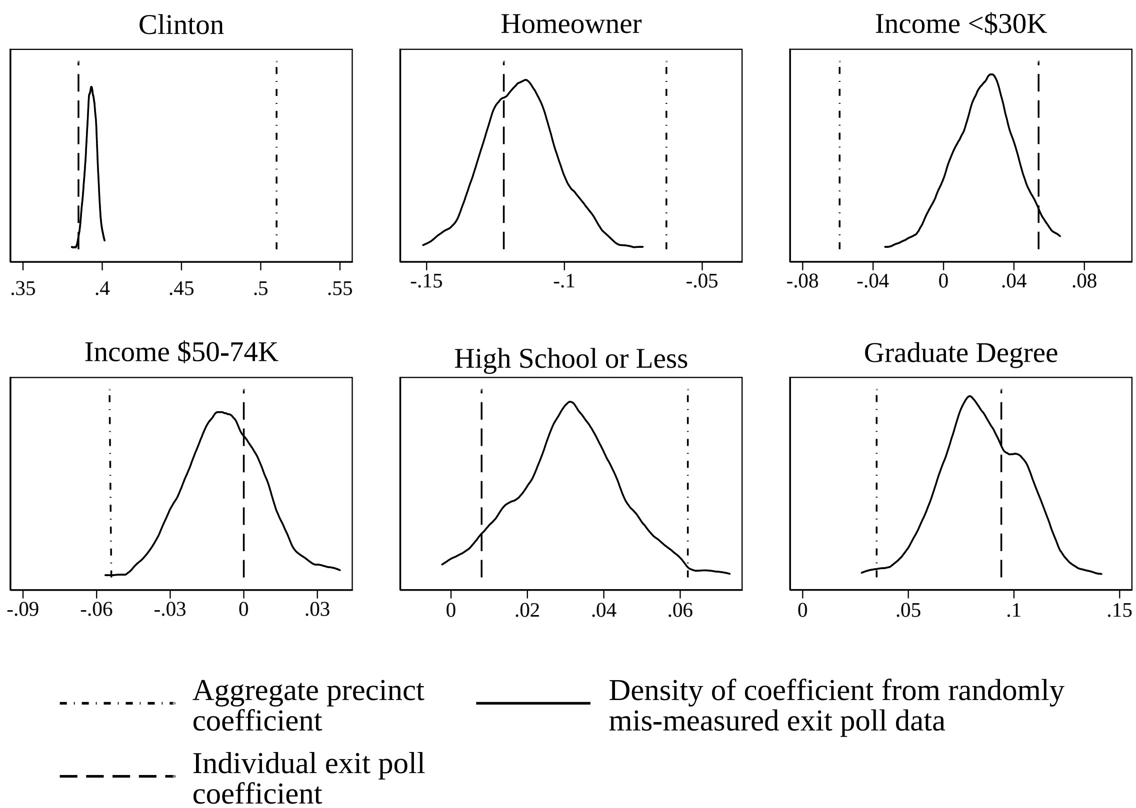
Notes: Results come from Monte Carlo simulation described in text. Vertical axis is the difference between unbiased estimate and the estimate derived from the aggregate model. The left plot sets $\mu = 0.2$, and the right plot sets $\mu = 0.8$. Box plot boundaries indicate 25th-75th percentile with the middle line being the median and the whiskers are the 5th and 95th percentiles. The mean of the unbiased estimates is 0.176. Other parameters are set as follows: $N = 250{,}000$, $J = 500$, $N_j = 500$, $\pi = 0.5$, $\beta = 1$, $\sigma = 2$, and $\omega \sim 80\%$.

Figure 4: Effect of selection into exit poll participation on coefficient bias



Notes: Results come from Monte Carlo simulation described in text. Vertical axis is the difference between the unbiased estimate and the estimate derived from the exit poll model. Horizontal axis is different values of θ_2, with $\theta_2 > 0$ implying $x = 1$ more likely to participate in the exit poll. Box plot boundaries indicate 25th-75th percentile with the middle line being the median and the whiskers are the 5th and 95th percentiles. The mean of the unbiased estimates is 0.176. Other parameters are set as follows: $N = 250{,}000$, $J = 500$, $N_j = 500$, $\pi_1 = 0.5$, $\pi_2 = 0.05$, $\theta_1 = 0$, $\beta = 1$, $\sigma = 2$, and $\omega \sim 95\%$.

Figure 5: Distribution of individual exit poll estimated coefficients when socioeconomic characteristics are randomly re-assigned to match population proportions



Notes: Socioeconomic characteristics of exit poll participants are randomly and iteratively changed from groups that are disproportionately more likely to vote relative to their population to groups that are disproportionately less likely to vote. Main specification is estimated on re-assigned sample and coefficients recorded. Process is repeated 200 times to estimate distribution of coefficients with mismeasured population characteristics. Aggregate precinct and individual exit poll coefficients come directly from Table 2.

Table 1: Summary statistics for aggregate data and exit poll

Variables (%)	Aggregate Precincts: Whole State (1)	Aggregate Precincts: Sampled Precincts (2)	Individual Exit Poll: Means (3)	Individual Exit Poll: Difference in means (4) = (3) - (2)	Individual Exit Poll: Within-precinct variation (%) (5)
Voted in favor of Green Economy Bonds (GEB)	67.6	66.3	66.3	0.0	92.7
Voted for Hillary Clinton	54.1	52.7	52.7	0.0	90.6
Voted for Donald Trump	39.2	40.3	40.3	0.0	90.6
Voted for Gary Johnson	3.2	3.4	3.4	0.0	98.3
Voted for Jill Stein	1.3	1.4	1.4	0.0	98.6
Voted for other presidential candidate	2.2	2.2	2.2	0.0	98.9
Homeowner	67.3	70.2	73.2	3	86.3
Female	52.2	51.1	52.6	1.5	97.2
Hispanic or Black	18.8	15.8	10.7	-5.1***	75.6
Age 18 to 29	19.3	18.2	13.3	-4.9***	95.8
Age 30 to 44	22.0	21.2	21.9	0.7	96.6
Age 45 to 54	18.4	19.2	22.5	3.3***	96.1
Age 55 to 64	18.4	18.9	22.3	3.4***	97.7
Age 65 or over	21.9	22.5	20.0	-2.5**	95.3
Household income less than $30K	22.3	21.0	12.2	-8.8***	91.7
Household income $30-49K	14.9	14.8	15.8	1	97.7
Household income $50-74K	16.0	14.7	21.6	6.9***	97.3
Household income $75-99K	13.5	13.7	16.7	3**	97.6
Household income over $100K	33.3	35.8	33.7	-2.1	89.7
Education level is high school or less	37.2	35.1	13.3	-21.8***	96.3
Education level is some college	26.4	26.7	27.7	1	97.0
Education level is college graduate	21.5	22.7	34.4	11.7***	97.5
Education level is graduate degree	14.9	15.5	24.7	9.2***	95.0
Observations (Precincts, Individuals)	416	37	2033		

Notes: For Columns 1 and 2, weighted means are displayed with precincts weighted by the total number of GEB votes. For Column 3, weighted means are displayed with individuals weighted by their sampling weight, which is determined by GEB and presidential vote (described in the main text). Column 4 presents differences in means and statistical significance of differences (*** $p<0.01$, ** $p<0.05$, * $p<0.1$). Statistical significance is determined using margin of error calculations by the American Community Survey and the standard error from the survey mean. Column 5 is calculated as 1-R-squared, with the R-squared coming from a weighted regression of each variable on precinct fixed effects.

Table 2: Determinants of voting in favor of Green Economy Bonds using different data sources

Variables % of precinct (column 1) or binary (column 2)	Aggregate precincts (1)	Individual exit poll (2)
Voted for Hillary Clinton	0.510***	0.385***
	(0.021)	(0.031)
Voted for Gary Johnson	0.461**	0.032
	(0.179)	(0.063)
Voted for Jill Stein	1.211***	0.156*
	(0.252)	(0.090)
Voted for other presidential candidate	0.148	0.135*
	(0.189)	(0.074)
Homeowner	-0.063***	-0.122***
	(0.017)	(0.031)
Female	0.059**	0.033*
	(0.028)	(0.018)
Hispanic or Black	-0.007	-0.038
	(0.010)	(0.044)
Age 18 to 29	0.001	0.013
	(0.024)	(0.040)
Age 45 to 54	0.036	-0.019
	(0.042)	(0.031)
Age 55 to 64	0.046	-0.014
	(0.038)	(0.034)
Age 65 or over	-0.079**	-0.023
	(0.032)	(0.035)
Household income less than $30K	-0.059**	0.054
	(0.027)	(0.043)
Household income $30-49K	-0.046*	-0.020
	(0.027)	(0.044)
Household income $50-74K	-0.054**	-0.000
	(0.027)	(0.029)
Household income $75-99K	0.003	0.023
	(0.037)	(0.028)
Education level is high school or less	0.062**	0.008
	(0.029)	(0.040)
Education level is college graduate	0.073**	0.027
	(0.032)	(0.031)
Education level is graduate degree	0.035	0.094**
	(0.030)	(0.040)
R-squared	0.926	0.234
Observations (Precincts, Individuals)	416	2,033

Notes: For Column 1, the unit of observation is precinct, the dependent variable is the proportion voting in favor of GEB, and observations are weighted by the total number of GEB votes. For Column 2, the unit of observation is the individual, the dependent variable is a binary indicator of voting in favor of GEB, and observations are weighted by their sampling weight, which is determined by GEB and presidential vote (described in the main text). Both columns are estimated using OLS. Additional control variables (measured at the precinct level) not displayed are: population density, residential property tax rate, average house sales price, distance to existing bike path, open space acres previously preserved by state funds, and number of remediated brownfields. Robust standard errors are used in Column 1, and precinct-clustered standard errors are used in Column 2. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Additional specifications for aggregate model

Variables (% of precinct)	(1)	(2)	(3)	(4)	(5)
Voted for Hillary Clinton	0.528***	0.510***	0.539***	0.533***	0.514***
	(0.020)	(0.021)	(0.023)	(0.024)	(0.022)
Voted for Gary Johnson	0.382**	0.461**	0.613***	0.565***	0.566***
	(0.169)	(0.179)	(0.181)	(0.184)	(0.184)
Voted for Jill Stein	1.190***	1.211***	1.503***	1.446***	1.194***
	(0.253)	(0.252)	(0.271)	(0.276)	(0.259)
Voted for other presidential candidate	0.074	0.148	0.419**	0.431**	0.109
	(0.183)	(0.189)	(0.172)	(0.172)	(0.193)
Homeowner	-0.055***	-0.063***	-0.041***	-0.042***	-0.053***
	(0.016)	(0.017)	(0.015)	(0.015)	(0.016)
Female	0.066**	0.059**	0.029	0.028	0.050
	(0.027)	(0.028)	(0.027)	(0.026)	(0.032)
Hispanic or Black	-0.006	-0.007	-0.004	-0.004	-0.006
	(0.010)	(0.010)	(0.012)	(0.012)	(0.018)
Age 18 to 29	-0.008	0.001	0.018	0.017	0.004
	(0.023)	(0.024)	(0.023)	(0.023)	(0.031)
Age 45 to 54	0.013	0.036	0.034	0.030	0.016
	(0.043)	(0.042)	(0.040)	(0.039)	(0.034)
Age 55 to 64	0.015	0.046	0.063	0.066*	0.024
	(0.036)	(0.038)	(0.038)	(0.038)	(0.030)
Age 65 or over	-0.103***	-0.079**	-0.050*	-0.050*	-0.078**
	(0.028)	(0.032)	(0.029)	(0.029)	(0.034)
Household income less than $30K	-0.044*	-0.059**	-0.011	-0.012	-0.113**
	(0.026)	(0.027)	(0.025)	(0.026)	(0.048)
Household income $30-49K	-0.033	-0.046*	-0.029	-0.024	-0.026
	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)
Household income $50-74K	-0.040	-0.054**	-0.010	-0.004	-0.052***
	(0.027)	(0.027)	(0.026)	(0.027)	(0.019)
Household income $75-99K	0.019	0.003	-0.011	-0.015	-0.020
	(0.035)	(0.037)	(0.028)	(0.028)	(0.027)
Education level is high school or less	0.065**	0.062**	0.019	0.014	0.169**
	(0.029)	(0.029)	(0.027)	(0.027)	(0.083)
Education level is college graduate	0.065**	0.073**	0.027	0.025	0.053***
	(0.031)	(0.032)	(0.031)	(0.031)	(0.020)
Education level is graduate degree	0.030	0.035	0.015	0.016	0.015
	(0.030)	(0.030)	(0.030)	(0.031)	(0.020)
Precinct characteristics		yes	yes	yes	yes
Town fixed effects			yes	yes	
Latitude/Longitude controls				yes	
Adjusted Demographics					yes
R-squared	0.924	0.926	0.949	0.950	0.925

Notes: For all columns, the unit of observation is precinct, the dependent variable is the proportion voting in favor of GEB, observations are weighted by the total number of GEB votes, models are estimated using OLS, and the sample size is 416. Precinct characteristics are: population density, residential property tax rate, average house sales price, distance to existing bike path, open space acres previously preserved by state funds, and number of remediated brownfields. Latitude/Longitude controls are quadratic functions of latitude and longitude and the interaction of latitude and longitude. For Column 5, census demographics are adjusted proportionately for each precinct so that the aggregate sample means are identical to those of the exit poll. Robust standard errors are used. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Additional specifications for individual model

Variables	(1)	(2)	(3)	(4)
Voted for Hillary Clinton	0.397***	0.385***	0.391***	0.390***
	(0.031)	(0.031)	(0.031)	(0.031)
Voted for Gary Johnson	0.035	0.032	0.043	0.041
	(0.061)	(0.063)	(0.063)	(0.063)
Voted for Jill Stein	0.164*	0.156*	0.168*	0.166*
	(0.091)	(0.090)	(0.095)	(0.095)
Voted for other presidential candidate	0.142*	0.135*	0.129	0.130
	(0.075)	(0.074)	(0.077)	(0.077)
Homeowner	-0.134***	-0.122***	-0.111***	-0.109***
	(0.031)	(0.031)	(0.031)	(0.032)
Female	0.031*	0.033*	0.031	0.030
	(0.018)	(0.018)	(0.019)	(0.019)
Hispanic or Black	-0.014	-0.038	-0.034	-0.038
	(0.039)	(0.044)	(0.045)	(0.047)
Age 18 to 29	0.006	0.013	0.007	0.009
	(0.039)	(0.040)	(0.041)	(0.041)
Age 45 to 54	-0.023	-0.019	-0.016	-0.017
	(0.032)	(0.031)	(0.031)	(0.031)
Age 55 to 64	-0.019	-0.014	-0.016	-0.016
	(0.034)	(0.034)	(0.035)	(0.035)
Age 65 or over	-0.028	-0.023	-0.020	-0.021
	(0.035)	(0.035)	(0.036)	(0.036)
Household income less than $30K	0.064	0.054	0.058	0.056
	(0.040)	(0.043)	(0.044)	(0.044)
Household income $30-49K	-0.014	-0.020	-0.007	-0.009
	(0.044)	(0.044)	(0.044)	(0.044)
Household income $50-74K	0.006	-0.000	0.002	0.000
	(0.029)	(0.029)	(0.030)	(0.030)
Household income $75-99K	0.030	0.023	0.025	0.023
	(0.029)	(0.028)	(0.030)	(0.030)
Education level is high school or less	0.006	0.008	0.004	0.005
	(0.040)	(0.040)	(0.041)	(0.041)
Education level is college graduate	0.026	0.027	0.022	0.020
	(0.030)	(0.031)	(0.031)	(0.031)
Education level is graduate degree	0.092**	0.094**	0.089**	0.088**
	(0.040)	(0.040)	(0.040)	(0.040)
Precinct characteristics		yes	yes	
Town fixed effects			yes	
Precinct fixed effects				yes
R-squared	0.227	0.234	0.254	0.257

Notes: For all columns, the unit of observation is the individual, the sample size is 2033, the dependent variable is a binary indicator of voting in favor of GEB, observations are weighted by their sampling weight, which is determined by GEB and presidential vote (described in the main text), and models are estimated using OLS. Precinct characteristics are: population density, residential property tax rate, average house sales price, distance to existing bike path, open space acres previously preserved by state funds, and number of remediated brownfields. Standard errors are clustered at the precinct level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Estimated willingness to pay for GEB and the distribution of benefits

	Aggregate precincts	Individual exit poll
Estimated willingness to pay (WTP) ($)	848	323
	(425)	(77)
Deviation from average WTP for various groups		
Democrat	128%	79%
Republican	-151%	-88%
Hispanic or Black	-10%	2%
White	2%	0%
Female	12%	5%
Male	-13%	-6%
Age 29 or under	13%	32%
Age 65 or over	-54%	-17%
Education level is high school or less	6%	-10%
Education level is graduate degree	-6%	26%

Notes: Regression models that enter into WTP calculation are presented in Appendix Table A7. Standard errors of estimated WTP are calculated using the delta method and are shown in parentheses.